



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

AlGashaam, Faisal, Nguyen Thanh, Kien, Chandran, Vinod, & Banks, Jasmine

(2017)

Elliptical higher-order-spectra pericocular code.

IEEE Access. (In Press)

This file was downloaded from: <https://eprints.qut.edu.au/106843/>

© 2017 The Author(s)

This work is licensed under a Creative Commons Attribution 3.0 License.
For more information, see <http://creativecommons.org/licenses/by/3.0>

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<https://doi.org/10.1109/ACCESS.2017.2697898>

Elliptical Higher-order-spectra Periocular Code

F. AlGashaam, *Member, IEEE*, K. Nguyen, *Member, IEEE*,
V. Chandran, *Senior Member, IEEE* and J. Banks, *Member, IEEE*

Abstract—The periocular region has recently emerged as a standalone biometric trait, promising attractive trade-off between the iris alone and the entire face, especially for cases where neither the iris nor a full facial image can be acquired. This advantage provides another dimension for implementing a robust biometric system, performed in non-ideal conditions. Global features (LBP, HOG) and local features (SIFT) have been introduced; however, the performance of these features can deteriorate for images captured in unconstrained and less-cooperative conditions. A particular set of Higher Order Spectral (HOS) features have been proved to be invariant to translation, scale, rotation, brightness level shift and contrast change. These properties are desirable in the periocular recognition problem to deal with the non-ideal imaging conditions. This paper investigates the HOS features in different configurations for the periocular recognition problem under non-ideal conditions. Especially, we introduce a new sampling approach for the periocular region based on an elliptical coordinate. This non-linear sampling approach is then combined with the robustness of the HOS features for encoding the periocular region. In addition, we also propose a new technique for combining left and right periocular. The proposed feature-level fusion approach bases on state-of-the-art bilinear pooling technique to allow efficient interaction between the features of both perioculars. We show the validity of the proposed approach in encoding discriminant features, outperforming or comparing favorably with the state-of-the-art features on the two popular datasets: FRGC and JAFFE.

Keywords: *non-ideal biometrics, periocular recognition, elliptical sampling, Higher Order Spectra, multi-instance fusion, multimodal compact bilinear pooling.*

I. INTRODUCTION

Biometrics has been shown to be critical to deal with the increasing incidents of fraud challenges in highly secure identity authentication systems. Unlike traditional token-based (e.g. cards, keys) and knowledge-based (e.g. PINs, passwords) approaches, biometrics cannot be lost, forgotten or shared. A number of human physiological and behavioral characteristics such as face, iris, fingerprint, keystroke, palm vein, retina and ear have been successfully used as biometrics [1]. Among these, iris and face are two of the most successful modalities that have widely been employed in security systems. However, in cases where the iris is occluded (e.g. eye lids, eye lashes, non-frontal gaze), of low resolution and quality due to the long distance from the

subject to the camera [2, 3] or the entire face cannot be obtained due to occlusion as shown in Figure 1; periocular - the region surrounding the eyes, may be visible and usable for recognition instead.

Periocular has emerged as a standalone biometric trait or complementary to the existing iris and face modalities [4]. The features proposed in the literature can be grouped in two groups: global features and local features. While the global features such as Local Binary Pattern (LBP) and Histogram of Gradient (HOG) [5] are extracted from the whole image or region of interest (ROI), the local features are extracted from a set of discrete points using such approaches as Scale Invariant Feature Transform (SIFT) [6] and Speeded Up Robust Features (SURF) [7]. The global features are usually in the forms of color, texture of the ROI and shape of eyelids and eyebrows. Even though the global and local features have been shown effective for the periocular recognition task, the quest for robust periocular features is still an active research area and an open question, especially when dealing with non-ideal and less cooperative imaging conditions such as deformation and noise in the periocular images [8].



Figure 1. When iris and face fail due to occlusion, low resolution and harsh illumination, periocular can be beneficial [9].

In this paper, we propose a novel approach for extracting periocular features. The proposed approach is based on a new sampling proposal and the use of Higher Order Spectra (HOS) to extract features from the sampled key points. The HOS technique has been long used for encoding discriminative features with advantages in its invariance properties against deformation [10]. Chandran and Elgar [10, 11] introduced a branch of 1 dimension (1D) HOS encoding techniques that is invariant to scale, level shift, translation and amplification. These properties are significant for dealing with the non-ideal deformation caused in the imaging systems. They also proposed a set of bispectral invariant features extracted from Radon

The paper is first submitted for review in 13rd March 2017.

Faisal AlGashaam conducts this research under the scholarship of the Ministry of Education, Saudi Arabian government 2014-2018.

F. AlGashaam, K. Nguyen, V. Chandran and J. Banks are with the Queensland University of Technology, 2 George Street, GPO Box 2434, Brisbane, Queensland 4001 (e-mail: faisal.algashaam@hdr.qut.edu.au, {k.nguyenthanh, v.chandran, j.banks}@qut.edu.au).

projections of images [10, 12] and a method to get rotation invariance. This paper investigates a variant of the technique applied to periocular recognition. Our major contribution is a novel approach to encode periocular based on HOS. The proposed approach is novel in decomposing the periocular region into an elliptical coordinate grid to make the encoding robust to scale, translation and simplifying head rotation, which may happen frequently in real life applications. This sampling technique coupled with the robustness of HOS features lead to a powerful feature extraction technique for periocular.

The presence of both left and right periocular in one captured image has been exploited for multi-instance fusion in the multi-biometric system [13]. Fusing multiple instances of the same body trait such as multiple fingers, left and right irises, and left and right periocular of the same individual has been shown effective. There are five levels of fusion: sensor-level, feature-level, score-level, rank-level and decision-level [14]. These levels perform the fusion in different stage of the recognition problem. While earlier stages retain more information for fusion, they also contain redundant information such as noise. Consequently, performing fusion at earlier stages (e.g. sensor-level) has more capability to play with the rich amount of information, but are more prone to noise and unexpected information. In contrast, later stages (e.g. rank-level and decision-level) have compact representation which is more related to the recognition process. Feature-level fusion is one of the most popular approaches in the literature due to its good trade-off between the richness and the compactness of representation. In this paper, we propose a new feature-level approach fusing two perioculars. While most state-of-the-art feature-level approaches base on sum, concatenation and element-wise multiplication, we approach the fusion from a completely new perspective based on bilinear pooling [15]. Our proposed multi-instance fusion approach for left and right periocular allows each element in both feature vectors interact with each other to find the correlation, leading to effective fusion.

This paper is an extension of our previous paper [16]. This paper introduces a new sampling technique for the periocular region to compensate for the imaging variations. In addition, we also propose a new fusion approach based on bilinear pooling to effectively fuse left and right perioculars of the same subject. These two new contributions result in higher accuracy of recognition.

The remaining of this paper is structured as follows: Section II reviews state-of-the-art approaches in periocular recognition; Section III provides technical background on Higher Order Spectra features; Section IV presents the proposed approach, Elliptical Higher-order-spectra Periocular Code, in recognizing perioculars; Section V explains a wide range of experiments on two databases: JAFFE and FRGC; and Section VI concludes the paper.

II. PERIOULAR RECOGNITION

The term “periocular” comes from the combination of two words---peri and ocular. While the former means the vicinity, the latter means something that is related to the eye [17, 18]. A periocular region is referred to as skin textures and anatomical features of the face region surrounding the eye, possibly including the eye, eyelids, eyelashes and eyebrows as shown in Figure 2. In 2009, Park et al. first investigated and proved that this region is sufficiently unique and reliable to serve as a standalone biometric [6, 16]. It can also be used complementary to iris and face biometrics, especially in unconstrained conditions when the full face cannot be obtained due to occlusion and the iris image is of poor quality due to its small size and short imaging distance constraint [4, 5].

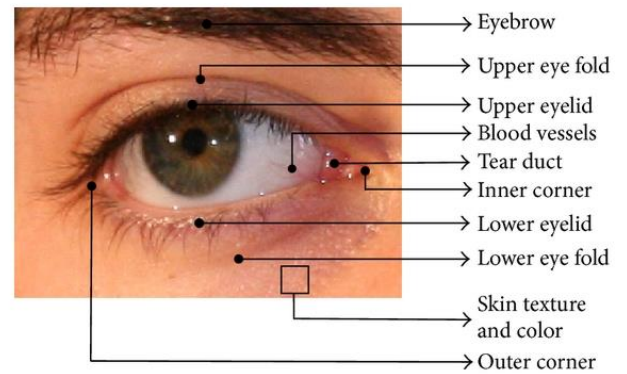


Figure 2. The areas surrounding the eyes that can be used for the periocular recognition task [5, 17].

Periocular recognition has emerged as one of the most active research areas in the biometrics. A typical periocular recognition approach first captured the periocular of the subject using a visible or near-infrared camera. The quality of the acquired image is checked. If the quality is acceptable with high resolution, sufficient illumination and blur free [18], the periocular region is then segmented from the image. The periocular region is normally a rectangular region localized by the eye center or the inner and outer corners of the eye. Choosing the best periocular region for recognition is still an open question for the research community [9]. The next step is extracting features. Choosing features representing reliable and discriminative properties of the periocular is one of the most critical tasks in the periocular recognition problem. This paper tackles this task and proposes a novel approach to efficiently extract discriminative properties of the periocular for high recognition accuracy. The final features are subsequently compared with the gallery features to find the match. The matching is performed through different classification techniques such as k Nearest Neighbor (k-NN), Support Vector Machine (SVM) and Gaussian Mixture Models (GMMs).

While LBP, HOG and SIFT features have been shown to be effective for the periocular recognition task [5, 17, 19],

other approaches have also been investigated. Bharadwaj *et al.* [2] proposed to combine global and local descriptor to better cover the variations of the inner class. The GIST descriptor was used as a global matcher, while LBP was used as the local matcher. The global scores and local scores are then normalized and combined by a weighted sum fusion approach. Experiments on the UBIRIS 2.0 dataset showed a rank-1 identification accuracy of 73.65%. Differently, Adams *et al.* [20] employed a genetic based Type II feature extraction approach to optimize the feature sets returned by LBP descriptors. The optimized feature sets significantly improved the recognition accuracy (10.99%) in comparison with the baselines. Subspace representation techniques have also been investigated by Juefei-Xu and Savvides [21] including Principal Component Analysis (PCA), Kernel Class-dependence Feature Analysis (KCFA), Unsupervised Discriminant Projection (UDP), and Kernel Discriminant Analysis (KDA). A significant boost in recognition performance was shown in comparison with the standard LBP descriptor and traditional subspace representations on raw pixel intensities.

Based on phase intensive patterns, Bakshi *et al.* introduced a novel multi-scale local feature of a visible spectrum periocular image. The periocular images are captured from subjects at a stand-off distance in less-cooperative scenarios [3]. The phase intensive patterns are extracted by a set of filters of different scales. While the number of features proposed is numerous, choosing the optimal feature is still open question for the research community. Xu *et al.* [22] tackled this question by comparing multiple features and found that the Local Walsh-Transform Binary Pattern outperformed all others, achieving the best accuracy, especially when fusing with Kernel Correlation Feature Analysis (KCFA) [23]. While most approaches are based on single image matching, Uzair *et al.* formulated the periocular recognition as an image set classification problem. They capitalized on the rich availability of multiple frames from the video in the Multi Biometric Grand Challenge to better deal with the periocular region variation [21]. The authors tested six state-of-the-art image set classification techniques and achieved significantly higher accuracy than single image approaches.

A number of work has studied the impact of non-ideal factors on the performance. Park *et al.* investigated the impact of pose, cosmetic changes, occlusion and template aging. Miller *et al.* in both [24] and [25] studied the non-ideal effect of blurring, image resolution and illumination conditions on the robustness of appearance-based periocular recognition. Their experiments showed that blurring caused a highest performance drop from 94.10% to 54.49% for neutral expressions compared across multiple sessions. In contrast, the resolution reduction does not severely downgrade the performance, decreasing from 94.90% to 84.70% when the images are down-sampled to 40%. To make the recognition robust to quality variation, Proença and Briceno in [26] studied a algorithm called Globally Coherent Elastic Graph Matching algorithm on the effect of

distortion. They showed improvement in recognition performance on the dataset FaceExpressUBI. Recently deep neural networks have been considered for periocular recognition [27, 28], but these approaches are a very computational demanding, normally require dedicated hardware such as Graphical Processing Units (GPUs).

Despite the rich literature on periocular recognition, there are few approaches dealing with scale change, translation, rotation, and illumination change. This is critical in non-ideal imaging conditions, which are more realistic in real life applications. Addressing these non-ideal conditions would further advance periocular biometrics and increase its applicability in broader range of applications.

III. HIGHER ORDER SPECTRAL FEATURES

Higher Order Spectra (HOS) were originally defined (in the 1960s) as spectral representations of cumulants of ergodic random processes [29-31] as a technique for characterizing non-Gaussian processes and nonlinear systems [32]. Subsequently they were introduced to signal processing and extended to apply to deterministic signals [33, 34]. Some of the earliest applications of HOS were to ocean waves [32] and EEG [35, 36]. One of the practical applications of HOS was the EEG Bispectral index for monitoring depth of anaesthesia [37, 38].

A set of invariant features for recognition of 1D patterns based on the bispectrum of a one-dimensional deterministic signal was introduced by Chandran and Elgar [11] and extended to apply to 2D images [10] using the Radon transform. These features have been adapted and applied to many real-world data sets and application domains such as sea-mine detection from sonograms [12], subsurface interface detection from GPR returns [39], virus classification from electron microscope images [40], speaker recognition [41], seizure detection from EEG [35] and breast cancer screening of thermogram images [42]. The feature extraction process is explained here.

If $x(n), n = 1, 2, \dots, N$ is a finite length 1D sequence with discrete-time Fourier transform $X(f)$, its bispectrum can be estimated as a function of two frequencies,

$$B(f_1, f_2) = E[X(f_1)X(f_2)^*(f_1 + f_2)] \quad ,$$

where the expectation operation can be removed if the signal is deterministic. It may be noted that in the presence of intra-class variations or noise, the 1D deterministic sequence will become a random process but stationarity and ergodicity assumptions are not required. The space of all estimate includes Dirac delta functions. In practice, the estimates are numerically computed using the discrete Fourier transform (DFT) after zero padding. Phases of integrals of the bispectrum along straight lines passing through the origin in the bi-frequency space were shown to be good candidates for robust features from 1D patterns in [10, 11]. Invariance properties were proved and profiles on bolts were classified with these features using a minimum distance classifier.

Later this method was extended and applied to 2D images using Radon projections in [10, 11]. Since the features from 1D patterns are designed to be invariant to the above transformations, the shift, scale and amplification invariance properties also transfer to 2D images. These features are invariant to a brightness shift (constant) affecting all pixels, a contrast stretch caused by illumination change and robust to small shifts and scale changes of the

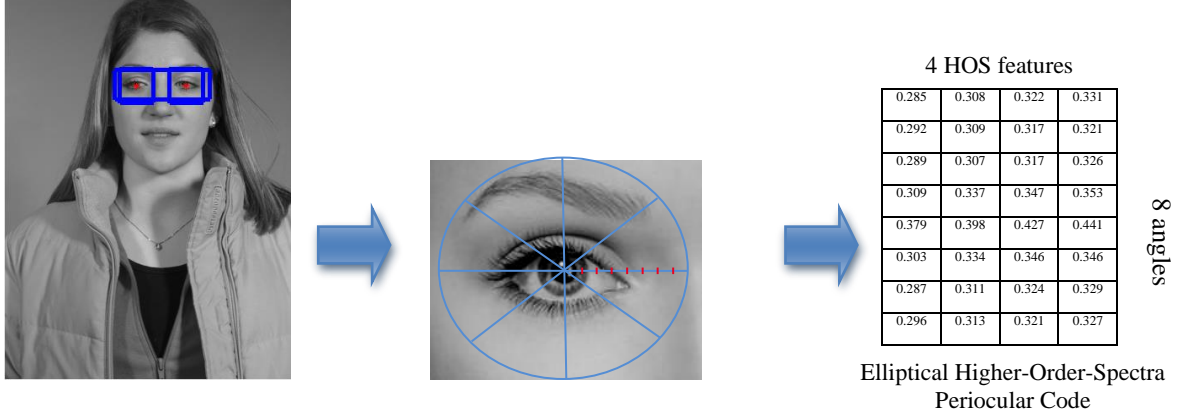


Figure 3. Illustrating the proposed approach with the image ID 04535d32 in the FRGC dataset. An elliptical coordinate is established inside the periocular region. The local points to encode are sampled at different angles and different radii within the elliptical coordinate. For each angle, the local point sequence is encoded by the HOS technique to generate 4 HOS features for each angle. The concatenation of HOS features for all angles (8 in this figure) makes the final periocular code called eHPC.

input image. Rotation invariance can be achieved by using the shift invariance property of a DFT taken along the angle or by finding the best correlation between test and reference feature sets for different angles as done with iris codes.

Despite the elegant invariance properties, this 2D HOS discussed above does not fit the periocular encoding problem since the Radon projects in the second step discard spatial information of the image, reducing the recognition performance. In this research, we propose a novel way to encode the periocular regions based on the HOS features. The strength of the proposed approach is achieved by both the invariance properties of the HOS features and the novel choice of the sampling coordinates for the periocular region.

IV. PROPOSED ELLIPTICAL HOS ENCODING APPROACH FOR PERIOULAR RECOGNITION

Suppose the head shot image in the dataset has a size of $H \times W$ pixels. For example, $H=256$, $W=256$ for the Japanese Female Facial Expression (JAFPE) dataset [43, 44]; $H=2272$, $W=1704$ for the Face Recognition Grand Challenge (FRGC) dataset. The periocular window size is $M \times N$. In the experiments, various window sizes will be tested to investigate how they affect the recognition performance. It is noteworthy that we employ the whole periocular region for feature extraction without masking the iris region.

In this research, we propose a new 2D HOS descriptor for the periocular region. A new sampling approach based on an elliptical coordinate surrounding the eyes will be introduced to deal with the rotation and scale variation in

the periocular region. The sampling approach extracts multiple 1D samples along the radius of the ellipse. The 1D HOS features are then calculated for each sample. The proposed elliptical HOS encoding approach is illustrated as in Figure 3. There are four steps of the proposed approach as described as follow:

- Firstly, the left and right eyes are detected and

located using Viola-Jones Haar Cascades [45] or Hough transform as described in [46].

- Secondly, if the distance from the subject to the camera varies (e.g. in the FRGC dataset), for each eye, its inner and outer corners are located. A Gabor eye-corner filter is constructed to detect the corner points as discussed in [47].
- Thirdly, an elliptical coordinate surrounding the eye is established. The horizontal axis of the ellipse is decided by the triple (eye center, inner corner and outer corner). The axis runs through the eye center and with orientation in parallel to the line between the inner and outer corners. The ellipse is centered at the eye center and is inscribed in the chosen window. The size of the window is a hyperparameter in this approach, which affects how much the surrounding areas to be incorporated in the periocular region. The local points are sampled at different angles and different radii. This research employed 8 angles (θ) and 8 radii (r).

-Lastly, for each sequence of local points (8 values at 8 radii) at each angle, encode them with the HOS invariant features [11]. We employed 4 HOS features for each sequence. Hence with 8 angles, 4 HOS features each angle, there are 8×4 features in total. This 8×4 matrix represents the encoded features of the periocular. We name the periocular feature matrix eHPC (Elliptical Higher-order-spectra Periocular Code). One example of this

eHPC for the image ID 04535d32 in the FRGC dataset is illustrated in Figure 3.

Matching: Euclidean distance is used for matching two eHPC feature vectors. The Euclidean distances are then fed into a k-NN classifier to find the match. As our approach focuses on feature extraction rather than classification, hence we want to hold the classifier simple and constant without extending to more complicated distances (such as Mahalanobis distance [48]) and more complicated classifiers (such as SVM [49] and Neural Network [50]). Using more complicated distances and classifiers may gain further accuracy for the proposed approach. We will prove that even with a simple classifier, we can achieve high recognition accuracy in comparison with other state-of-the-art approaches. In order to make sure that the conclusions are not dependent on one particular K used, we will vary K.

The properties of the proposed sampling and feature extraction technique are:

- The use of the elliptical coordinate makes a head rotation become a shift in the eHPC. The number of angles in the elliptical coordinate can be adjusted according to the expected head rotation angles. Smaller rotation angles can be handled by using a higher number of angles in the encoding step. Otherwise, there will be some interpolation error leading to a perturbed feature value.
- A method that seeks a trade-off in the amount of information of the periocular to use in terms of the number of angles and the number of local points for each angle.
- It takes advantages of invariance property of the HOS for each local point sequence in each angle. This makes the eHPC robust to rotation, scale, level shift, translation and amplification.
- The presence of facial expressions may severely deform the periocular, which in turn decreases the recognition accuracy. To deal with this, we set up the training with a number of pre-defined expressions to allow the system to flexibly deal with the distortion caused by facial expressions. For example, for the JAFFE dataset, we will the train data include 7 different facial expressions (neutral, happy, angry, disgusted, feared, sad and surprised).
- The iris texture is not explicitly used in this approach and neither is it excluded or masked. Other approaches such as [17] exclude the iris region by a mask. The same could be applied here. However, we want to focus on the invariance properties the proposed feature extraction; hence, we do not use the mask for the iris region in our approach.
- The weakness of the proposed approach is the distortion caused by an up-down movement of the face, which is not compensated by the elliptical coordinate proposed. This can be coped with by head pose estimation and normalizing the in plane

rotation as a pre-processing step. In this research, the in plane rotation does not exist in the two datasets we are working on; hence for simplicity, the normalization step is not mentioned.

Fusion: The left periocular and the right periocular can be used either individually or in combination. When combined, as known as multi-instance fusion in the multibiometric community [14], they can provide rich information to increase robustness and accuracy of the whole biometric system. We propose a feature-level fusion approach, which relies on Multimodal Compact Bilinear pooling (MCB) to get a joint representation of two perioculars. MCB pooling has recently been proposed to effectively combine two modalities: image features and textual features in the visual question answering task [51]. Bilinear pooling computes the outer product between two feature vectors, which allows, in contrast to element-wise product, a multiplicative interaction between all elements of both vectors [15]. This multiplicative interaction is especially useful for combining two features because it allows the system to learn the correlation between them. Despite this property, bilinear pooling has not been explored for multibiometrics due to its high dimensionality (n^2). Only recently, Fukui *et al.* [51] has brought this benefit to multi-modal fusion by randomly projecting the image and text representations in the visual question answering task to a higher dimensional space using Count Sketch [52] and then convolving both feature vectors efficiently by using element-wise product in Fast Fourier Transform (FFT) space as shown in Figure 4. In this paper, we adopt this Multimodal Compact Bilinear pooling approach to combine features of the left and right periocular of one subject.

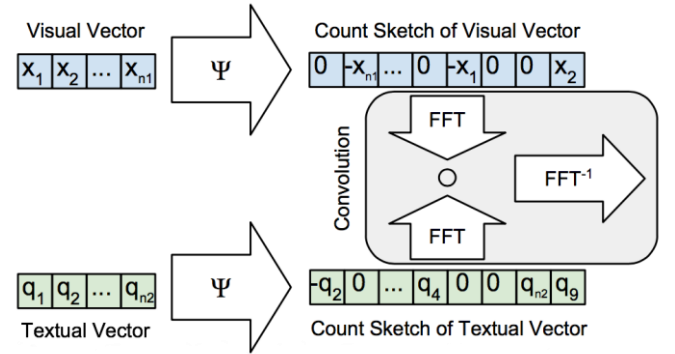


Figure 4. Multimodal Compact Bilinear Pooling [51]

V. EXPERIMENTAL RESULTS

In this research, we choose two datasets for experiments because of various facial expressions presence in them.

- The first one is the Japanese Female Facial Expression (**JAFFE**) database [43, 44]. This is a small dataset containing 213 images from 10 Japanese females with 7 facial expressions (NE: neutral, HA: happy, AN: angry, DI: disgusted, FE: feared, SA: sad, SU: surprised). The images are gray scale with a resolution of (256×256) pixels.

Figure 5 illustrates a set of seven emotional expressions of one subject. The dataset is divided into 2 subsets: TRAIN and TEST. For each subject, 7 images representing 7 various expressions are randomly selected for the TEST subset. The remaining images, which are approximately 15 images, are used for the TRAIN subset.

- The second one is the Face Recognition Grand Challenge 2.0 (**FRGC**) [53]. It contains frontal images of subjects captured in a studio setting, with controlled illumination and background. The resolution of images is (2272×1704) pixels. Two facial expressions (i.e. neutral and smiling) are recorded for each subject. Figure 6 depicts these two facial expressions of a subject. Three images (2 neutrals and 1 smiling) of each subject in 568 subjects are captured multiple times in various studio-imaging configurations, resulting in a total of 5451 face images. An example of emotional expression presented in the FRGC dataset is given in Figure 6. The dataset is divided into 2 subsets for these experiments: TRAIN and TEST. For each subject, four images of neutral expression and 4 images of smiling expression are randomly selected for the TRAIN subset. The remaining images are used for TEST.

The presence of various expressions in the periocular region is challenging for the recognition task, especially when using only one gallery image, leading to a very high computation complexity due to the non-linear deformation of the periocular regions caused by facial expression as shown in Figure 5. This challenge has also been recorded in other biometric modalities such as face recognition [54]. This challenge can be effectively dealt with by utilizing multiple gallery images with multiple expressions as in [55]. The rich availability of multi-expressions in the gallery helps to better deal with a wide range of periocular deformation at the cost of the computing expense.

It is noteworthy that the eye centers in the JAFFE dataset have been pre-defined at certain locations since the faces have been aligned. For the FRGC dataset, we employ the Viola-Jones algorithm [56] introduced in the Matlab Computer Vision Toolbox [57]. The detection results were verified visually in the pre-processing step.



Figure 5. Sample images from the JAFFE dataset with different expressions, which significantly affect the

periocular regions and the accuracy of a periocular recognition system [43, 44].



Figure 6. Sample images with different expressions and different studio imaging configurations from the FRGC dataset [53].

We now perform experiments varying parameter combinations for the classifiers and the window to investigate how the proposed approach performs over different scenarios.

1. The size of the periocular window

The size of the elliptical window decides how much surrounding detail is taken into account. A grid search with variation of the width and the height of the window is used to investigate their impacts. For this experiment, the k-NN classifier with the number of nearest neighbor of 3, the number of angles of 8 and the number of local points of 8 are chosen. Table 1 and 2 show that short and tall (in terms of the height) and narrow and wide (in terms of the width) have adverse impacts on the recognition performance. The best size of the window in terms of recognition accuracy is (M,N) = (38,34) pixels for the left periocular and (M,N) = (40,38), (42,36) and (42,38) for the right periocular in the JAFFE dataset. Similarly, the optimal window size is (38,34) for the left periocular and (42,32) and (42,34) for the right periocular in the FRGC dataset. It is interesting that the left and right periocular do not yield the same accuracy for both datasets. An insight in the datasets shows that for a number of persons, the left and right eyes differ, which results in the difference in the accuracy. For the rest of the experiments, these optimal window sizes are to be used if not other specified.

Table 1. Impact of the periocular window size to the accuracy (%) of LEFT periocular (JAFFE)

Height	Width					
	32	34	36	38	40	42
30	79.03	87.10	85.48	85.48	79.03	74.19
32	79.03	87.10	83.87	82.26	79.03	82.26
34	82.26	88.71	88.71	91.94	85.48	87.10
36	80.65	87.10	88.71	88.71	83.87	87.10
38	77.42	80.65	83.87	87.10	83.87	88.71
40	76.58	75.81	82.26	87.10	82.26	85.48

Table 2. Impact of the periocular window size to the accuracy (%) of RIGHT periocular (JAFPE)

Width						
Height	32	34	36	38	40	42
30	79.03	74.19	80.65	74.19	79.03	82.26
32	72.58	75.81	80.65	70.79	74.19	70.97
34	64.52	75.81	72.58	77.42	77.42	82.26
36	69.35	80.65	77.42	80.65	77.42	83.87
38	74.19	75.81	75.81	77.42	83.87	83.87
40	67.74	72.58	75.81	79.03	80.65	82.26

Table 3. Impact of the periocular window size to the accuracy (%) of LEFT periocular (FRGC)

Width						
Height	32	34	36	38	40	42
30	81.82	79.80	81.82	81.82	79.80	82.83
32	82.83	84.85	81.82	80.81	81.82	83.84
34	85.86	84.85	82.83	86.87	80.81	80.81
36	84.85	83.84	82.83	75.76	78.79	79.80
38	83.84	79.80	81.82	80.81	77.78	79.80
40	85.86	85.86	80.81	80.81	79.80	83.84

Table 4. Impact of the periocular window size to the accuracy (%) of RIGHT periocular (FRGC)

Width						
Height	32	34	36	38	40	42
30	87.88	90.91	92.93	93.94	91.92	92.93
32	90.91	88.89	89.90	92.93	93.94	94.95
34	87.88	83.84	88.89	87.88	92.93	94.95
36	87.88	88.89	86.87	87.88	90.91	92.93
38	89.90	89.90	88.89	88.89	87.88	88.89
40	89.90	90.91	87.88	87.88	83.84	91.92

2. The number of angles and radius local points

The number of angles and the number of local points chosen over each angle radius impact the details incorporated in the feature. The higher the numbers, the denser the sampling and the more detail incorporated. Choosing these number is balancing between detail amount and generalization. In the experiments with the window size of (38,34) pixels, we tested with 3 different numbers of local points along each radius, including 8, 12 and 16. In addition, we tested with 4 different numbers of angles, including 4, 8, 12 and 16. The highest accuracy achieved is 99.13% at 12 local points and 12 angles for the JAFPE dataset and 96.84% at 12 local points and 8 angles for the FRGC dataset.

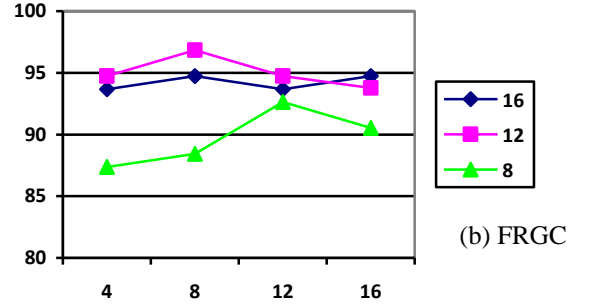
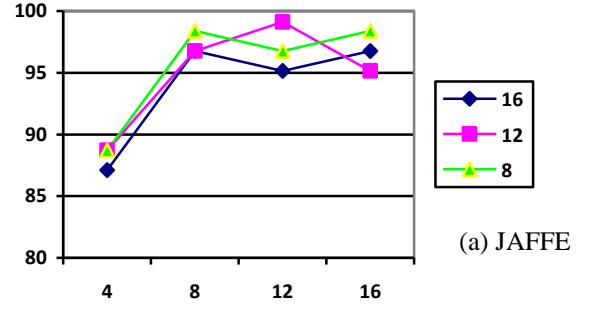


Figure 7. Accuracies (%) of the proposed approach with various values of angle and number of local points chosen over each angle radius on (a) JAFPE and (b) FRGC dataset.

3. The number of nearest neighbor K

The number of K nearest neighbor decides the number of samples considered, affecting the performance of the whole system in terms of the robustness to noise. Since there are seven images in the TRAIN for each subject in the JAFPE dataset and eight images for each subject in the FRGC dataset, we tested odd values from 1 to 7 for K. Increasing the value of K from 1 to 3 improves the recognition accuracy for both left periocular and fusion approaches. However, the system drops the accuracy when the value of K is greater than three. The results are presented in Figure 8.

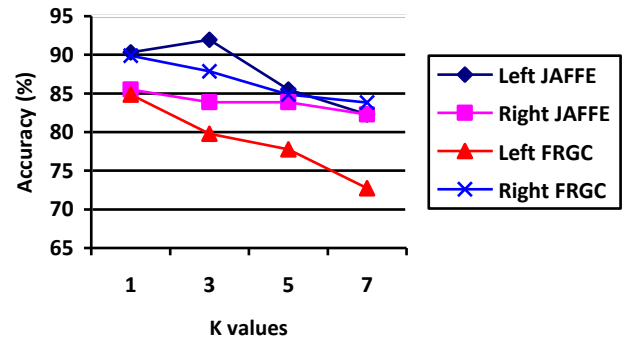


Figure 8. Accuracies (%) of the proposed approach with various values of K nearest neighbor.

4. Multi-instance fusion of left periocular and right periocular

Features are separately extracted from both perioculars then combined to generate a final feature vector. This combined feature vector is then used to compared with the TRAIN combined feature vectors to calculate the accuracy for feature-level fusion. Besides implementing the proposed multimodal compact bilinear pooling approach, we also implemented other feature-level fusion approaches including concatenation and element-wise multiplication for comparison. In addition, a score-level fusion approach using weighted sum was also implemented to compare with the proposed feature-level approach. The image-level fusion is indirectly inferred since the elliptical coordinate using for calculating features is wider in the horizontal axis than in the separate periocular case. The performance of fusion is presented in Table 5. The results show the proposed feature-level fusion approach outperforms both image-level and score-level fusions. This is due to the good trade-off of the feature-level fusion between the amount of information to retain and the robustness of the representation. The proposed multimodal bilinear pooling approach also outperforms other feature-level approach, showing its effectiveness in modeling multiplicative interaction between two feature vectors of two periocular. The improvement of the accuracy demonstrates the benefit of fusion.

The genuine and impostor score distributions of feature-fusion approach for the combined dataset (combining JAFFE and FRGC) are illustrated in Figure 9. The figure shows that the genuine scores and impostor scores are well separated, leading to a high recognition accuracy of the proposed approach.

Table 5: Accuracies (%) of the proposed approach with different fusions.

	JAFFE	FRGC
Image-level Fusion	96.77	88.89
Score-level Fusion	93.55	90.91
Feature-level Fusion		
Concatenation	97.23	94.71
Element-wise multiplication	98.39	95.96
Ours	99.52	97.71

5. Comparison with other state-of-the-art approaches

We also compare the proposed approach with the state-of-the-art approaches. The most popular choices are LBP, HOG and SIFT using the VLFeat open source library [58]. Following [5], we quantize both HOG and LBP into 8 distinct values to construct an eight-bin histogram. This histogram is generated from a partitioned sub-region and concatenated to output the final feature vector. The HOG is implemented with a cell size of 8. The SIFT is implemented with the peak threshold of 20.0 and the edge threshold of 5.0 to keep the dimension of the descriptor reasonable. The properties of these descriptors may vary with different window sizes. To be fair, we compare with them at different window size. The accuracy is presented in Table 6 for the JAFFE dataset and Table 7 for the FRGC dataset. The proposed approach outperforms all three LBP, HOG and SIFT approaches in terms of the accuracy.

Table 6. Comparison of the proposed approach and other state-of-the-art approaches on the JAFFE dataset.

Window size	LBP	HOG	SIFT	HOS
34,34	90.32	88.71	75.81	93.55
34,36	91.94	87.10	67.74	91.94
34,38	90.32	85.48	75.81	91.94
34,40	93.55	91.94	66.13	88.71
36,34	95.16	91.94	70.97	93.55
36,36	95.16	91.94	62.90	91.94
36,38	90.32	87.10	77.42	90.32
36,40	91.94	88.71	69.35	93.55
38,34	95.16	90.32	62.90	93.55
38,36	93.55	88.71	61.29	96.77
38,38	90.32	88.71	82.26	93.55
38,40	90.32	85.48	75.81	94.94
40,34	91.94	88.71	66.13	95.16
40,36	87.10	85.48	61.29	95.16
40,38	91.94	83.87	77.42	98.39
40,40	90.32	85.48	69.35	94.94
42,34	95.16	87.10	67.74	96.77
42,36	83.87	88.71	70.97	95.16
42,38	91.94	82.26	72.58	90.32
42,40	90.32	83.87	72.58	91.94

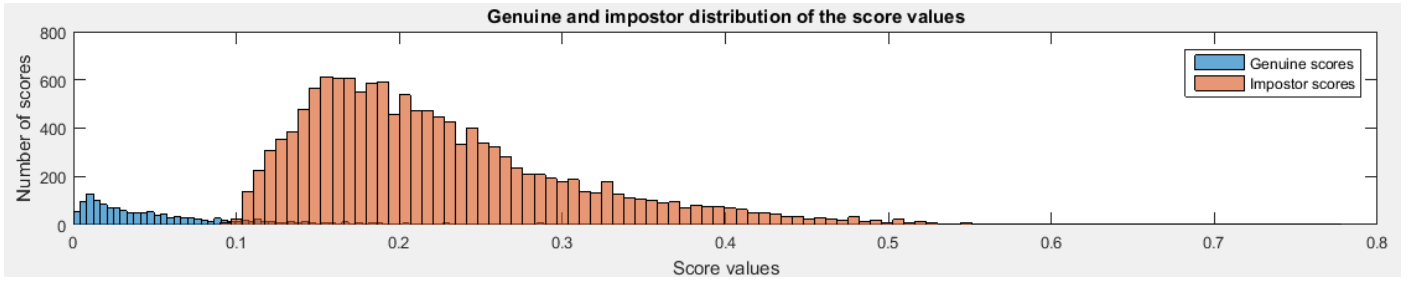


Figure 9. The genuine and impostor score distribution of the combined dataset (JAFEE and FRGC).

Table 7. Comparison of the proposed approach and other state-of-the-art approaches on the FRGC dataset.

Window size	LBP	HOG	SIFT	HOS
34,34	87.37	89.90	67.68	93.94
34,36	88.42	88.89	54.55	94.95
34,38	89.47	88.89	62.63	93.94
34,40	88.42	91.92	58.59	93.94
36,34	87.37	89.90	55.56	92.93
36,36	89.47	88.89	62.63	92.93
36,38	89.47	88.89	62.63	94.95
36,40	89.47	92.93	66.67	94.95
38,34	88.42	89.90	74.75	95.96
38,36	91.58	88.89	51.52	93.94
38,38	89.47	88.89	56.57	92.93
38,40	89.47	91.92	50.51	94.95
40,34	91.58	88.89	77.78	94.95
40,36	90.53	88.89	65.66	95.96
40,38	89.47	88.89	65.66	92.93
40,40	90.53	91.92	63.64	92.93
42,34	91.58	90.91	79.80	95.96
42,36	90.53	88.89	65.66	95.96
42,38	90.53	90.91	67.68	94.95
42,40	91.58	94.95	63.64	94.95

In the literature, the FRGC dataset has been widely used to publish the results by other researchers. Our results outperform the reported state-of-the-art in this dataset as shown in Table 8.

Table 8. Comparison of the proposed approach with other state-of-the-art approaches reported on the FRGC dataset.

[24, 25]	[20]	[59, 60]	Ours
89.1	92.2	91.0	95.96

CONCLUSION

Periocular is one of the emerging biometric modalities, which can function either independently as a standalone biometric modality or complementary to face and iris modalities. In this research, we have proposed a novel encoding technique for periocular. The proposed technique combines the invariance properties of 1D Higher Order Spectral features with an elliptical coordinate sampling technique to achieve robustness in scale, translation and head rotation. We have extensively investigated different configurations of hyper-parameters in the proposed approach to analyze its performance. It consistently outperforms the baselines (LBP, HOG and SIFT) on both JAFEE and FRGC datasets. The proposed approach achieves state-of-the-art recognition accuracy in these datasets. In addition, we also propose a new feature-level fusion approach for combining the left and right periocular. This fusion approach relies on multimodal compact bilinear pooling to enable multiplicative interaction for efficient correlation modeling between two feature vectors. This has been shown to outperform other level fusion approaches as well as other feature-level approaches. In the future work, we will experiment our proposed eHPC for periocular recognition in large-scale datasets such as LFW and MegaFace with millions of face images.

REFERENCE

- [1] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4-20, 2004.
- [2] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *IEEE International Conference BTAS*, 2010, pp. 1-6.
- [3] S. Bakshi, P. K. Sa, and B. Majhi, "A novel phase-intensive local pattern for periocular recognition under visible spectrum," *BIOCYBERNETICS AND BIOMEDICAL ENGINEERING*, vol. 35, pp. 30-44, 2015.
- [4] I. Nigam, M. Vatsa, and R. Singh, "Ocular biometrics: A survey of modalities and fusion approaches," *Information Fusion*, vol. 26, pp. 1-35, 2015.

- [5] P. Unsang, R. Jillela, A. Ross, and A. K. Jain, "Periocular Biometrics in the Visible Spectrum," *Information Forensics and Security, IEEE Transactions on*, vol. 6, pp. 96-106, 2011.
- [6] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, pp. 1150-1157 vol.2.
- [7] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Computer Vision – ECCV 2006*. vol. 3951, A. Leonardis, H. Bischof, and A. Pinz, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 404-417.
- [8] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," in *Biometrics (ICB), 2012 5th IAPR International Conference on*, 2012, pp. 439-445.
- [9] J. M. Smereka and B. V. K. V. Kumar, "What Is a "Good" Periocular Region for Recognition?," in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 117-124.
- [10] V. Chandran, B. Carswell, B. Boashash, and S. Elgar, "Pattern recognition using invariants defined from higher order spectra: 2-D image inputs," *Image Processing, IEEE Transactions on*, vol. 6, pp. 703-712, 1997.
- [11] V. Chandran and S. L. Elgar, "Pattern Recognition Using Invariants Defined From Higher Order Spectra- One Dimensional Inputs," *Signal Processing, IEEE Transactions on*, vol. 41, p. 205, 1993.
- [12] V. Chandran, S. Elgar, and A. Nguyen, "Detection of mines in acoustic images using higher order spectral features," *IEEE Journal of Oceanic Engineering*, vol. 27, pp. 610-618, 2002.
- [13] A. a. J. Ross, Anil K and Nandakumar, Karthik, "Information Fusion in Biometrics," in *Handbook of multibiometrics*. vol. 6, ed, 2006, pp. 37-58.
- [14] A. Ross, A. Jain, and K. Nandakumar, "Levels of Fusion in Biometrics," in *Handbook of Multibiometrics*, ed, 2006, pp. 59-90.
- [15] T. Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1449-1457.
- [16] F. AlGashaam, N. Kien, W. Boles, and V. Chandran, "Periocular recognition under expression variation using Higher Order Spectral features," in *2015 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2015, pp. 1-6.
- [17] P. Unsang, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *IEEE International Conference on BTAS*, 2009, pp. 1-6.
- [18] C. W. Tan and A. Kumar, "Towards Online Iris and Periocular Recognition Under Relaxed Imaging Constraints," *IEEE Transactions on Image Processing*, vol. 22, pp. 3751-3765, 2013.
- [19] G. Mahalingam, K. Ricanek, and A. M. Albert, "Investigating the Periocular-Based Face Recognition Across Gender Transformation," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 2180-2192, 2014.
- [20] J. Adams, D. L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn, "Genetic-Based Type II Feature Extraction for Periocular Biometric Recognition: Less is More," in *International Conference on Pattern Recognition (ICPR)*, pp. 205-208.
- [21] F. Juefei-Xu and M. Savvides, "Subspace-Based Discrete Transform Encoded Local Binary Patterns Representations for Robust Periocular Matching on NIST's Face Recognition Grand Challenge," *IEEE Transactions on Image Processing*, vol. 23, pp. 3490-3505, 2014.
- [22] J. Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *IEEE International Conference BTAS*, 2010, pp. 1-8.
- [23] C. Xie, M. Savvides, and B. V. K. VijayaKumar, "Kernel Correlation Filter Based Redundant Class-Dependence Feature Analysis (KCFA) on FRGC2.0 Data," in *Analysis and Modelling of Faces and Gestures: Second International Workshop, AMFG 2005, Beijing, China, October 16, 2005. Proceedings*, W. Zhao, S. Gong, and X. Tang, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 32-43.
- [24] P. Miller, A. Rawls, S. Pundlik, and D. Woodard, "Personal identification using periocular skin texture," in *ACM Symposium on Applied Computing*, 2010, pp. 1496-1500.
- [25] P. E. Miller, J. R. Lyle, S. J. Pundlik, and D. L. Woodard, "Performance evaluation of local appearance based periocular recognition," in *IEEE International Conference BTAS*, 2010, pp. 1-6.
- [26] H. Proença and J. C. Briceño, "Periocular biometrics: constraining the elastic graph matching algorithm to biologically plausible distortions," *IET Biometrics*, vol. 3, pp. 167-175, 2014.
- [27] R. Raghavendra and C. Busch, "Learning deeply coupled autoencoders for smartphone based robust periocular verification," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 325-329.
- [28] Z. Zhao and A. Kumar, "Accurate Periocular Recognition Under Less Constrained Environment Using Semantics-Assisted Convolutional Neural Network," *IEEE Transactions on Information*

- Forensics and Security*, vol. 12, pp. 1017-1030, 2017.
- [29] D. R. Brillinger, "An Introduction to Polyspectra," *The Annals of Mathematical Statistics*, vol. 36, pp. 1351-1374, 1965.
- [30] D. R. Brillinger and M. Rosenblatt, *Asymptotic theory of estimates of k-th order spectra*: New York: Wiley, 1967.
- [31] D. R. Brillinger and M. Rosenblatt, *Computation and Interpretation of k-th order spectra*: New York: Wiley, 1967.
- [32] H. K. Hasselmann, W. H. Munk, and G. J. F. MacDonald, *Bispectra of ocean waves*: New York: Wiley, 1963.
- [33] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE*, vol. 75, pp. 869-891, 1987.
- [34] C. L. Nikias and A. Petropulu, *Higher-order spectra analysis: A nonlinear signal processing framework*: NJ: Prentice Hall, 1993.
- [35] K. C. Chua, V. Chandran, R. Acharya, and C. M. Lim, "Higher Order Spectral (HOS) Analysis Of Epileptic EEG Signals," United States, pp. 6495-6498.
- [36] G. Dumermuth, P. J. Huber, B. Kleiner, and T. Gasser, "Analysis of the interrelations between frequency bands of the EEG by means of the bispectrum a preliminary study," *Electroencephalography and Clinical Neurophysiology*, vol. 31, pp. 137-148, 1971.
- [37] M. S. Avidan, L. Zhang, B. A. Burnside, K. J. Finkel, A. C. Searleman, J. A. Selvidge, *et al.*, "Anesthesia Awareness and the Bispectral Index," *The New England Journal of Medicine*, vol. 358, pp. 1097-1108, 2008.
- [38] S. D. Greenwald, C. P. Smith, J. C. Sigl, H. M. Cai, and P. H. Devlin, "The EEG Bispectral Index™/ (BISTM): development and utility," 1999, p. 443 vol.1.
- [39] V. Chandran, D. Ning, and S. Sridharan, "Speaker identification using higher order spectral phase features and their effectiveness vis-a-vis Mel-Cepstral features," in *Biometric Authentication. First International Conference, ICBA 2004. Proceedings, 15-17 July 2004*, Berlin, Germany, 2004, pp. 614-22.
- [40] A. D. Strange, V. Chandran, and J. C. Ralston, "Subsurface Interface Detection Using Bispectral Features and Ground Penetrating Radar."
- [41] H. Ong and V. Chandran, "Identification of gastroenteric viruses by electron microscopy using higher order spectral features," *Journal of Clinical Virology*, vol. 34, pp. 195-206, 2005.
- [42] M. EtehadTavakol, V. Chandran, E. Y. K. Ng, and R. Kafieh, "Breast cancer detection from thermal images using bispectral invariant features," *International Journal of Thermal Sciences*, vol. 69, pp. 21-36, 2013.
- [43] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, pp. 1357-1362, 1999.
- [44] M. N. Dailey, C. Joyce, M. J. Lyons, M. Kamachi, H. Ishi, J. Gyoba, *et al.*, "Evidence and a Computational Explanation of Cultural Differences in Facial Expression Recognition," *Emotion*, vol. 10(6), pp. 874-893, 2010.
- [45] P. Viola and M. Jones, "Robust Real-time Object Detection," *International Journal of Computer Vision*, 2001.
- [46] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 21-30, 2004.
- [47] Z. Zheng, J. Yang, and L. Yang, "A robust method for eye features extraction on color image," *Pattern Recognition Letters*, vol. 26, pp. 2252-2261, 2005.
- [48] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognition*, vol. 41, pp. 3600-3612, 2008.
- [49] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [50] B. M. Ozyildirim and M. Avci, "Generalized classifier neural network," *Neural networks : the official journal of the International Neural Network Society*, vol. 39, pp. 18-26, 2013.
- [51] A. Fukui, P. Dong Huk, Y. Daylen, A. Rohrbach, T. Darrell, and M. Rohrbach, "Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding [arXiv]," *arXiv*, p. 13 pp., 06/06 2016.
- [52] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact Bilinear pooling," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, June 26, 2016 - July 1, 2016*, Las Vegas, NV, United states, 2016, pp. 317-326.
- [53] NIST. *Face Recognition Grand Challenge Database* [Online]. Available: <http://www.nist.gov/itl/iad/ig/frgc.cfm>
- [54] L. Zhang, D. Tjondronegoro, V. Chandran, and J. Eggink, "Towards robust automatic affective classification of images using facial expressions for practical applications," *Multimedia Tools and Applications*, pp. 1-27, 2015.
- [55] M. Uzair, A. Mahmood, A. Mian, and C. McDonald, "Periocular biometric recognition using image sets," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, 2013, pp. 246-251.
- [56] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in

Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, pp. I-511-I-518 vol.1.

- [57] I. The MathWorks, "MATLAB and Computer Vision Toolbox Release 2015a," ed. USA, 2015.
- [58] A. Vedaldi and B. Fulkerson. (2008). *VLFeat: An Open and Portable Library of Computer Vision Algorithms*. Available: <http://www.vlfeat.org/>
- [59] D. L. Woodard, S. J. Pundlik, J. R. Lyle, and P. E. Miller, "Periocular region appearance cues for biometric identification," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, pp. 162-169.
- [60] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the Fusion of Periocular and Iris Biometrics in Non-ideal Imagery," in *20th International Conference on Pattern Recognition (ICPR)*, 2010, pp. 201-204.